Analysis of Genetic Interaction Maps Reveals Functional Pleiotropy

Shuye Pu[1], Karen Ronen [1,] James Vlasblom[1], Jack Greenblatt[2,3], Shoshana J. Wodak[1,3,4*]

[1]*Molecular structure and Function program, Hospital for Sick Children, 555 University Avenue, Toronto ON, Canada M5G 1X8.* [2]*Terrence Donnelly Centre for Cellular & Biomolecular Research (Donnelly CCBR), 160 College Street, Room 930 Toronto, Ontario Canada, ON M5S 3E1.* [3]*Department of Molecular Genetics and* [4]*Department of Biochemistry, University of Toronto, 1 Kings College Circle, Toronto ON, Canada M5S 1A8*

*To whom correspondence should be addressed: shoshana@sickkids.ca

1. INTRODUCTION

Epistatic or genetic interactions, representing the effects of mutations on the phenotypes caused by other mutations, can be very helpful for uncovering functional relationships between genes. Recently, the Epistasis Miniarray Profile (E-MAP) method has emerged as a powerful approach for identifying such interactions systematically. As part of this approach, hierarchical clustering is used to partition genes into groups on the basis of the similarity between their global interaction profiles. Here we present an original biclustering algorithm, termed Local Coherence Detection (LCD) algorithm, which exploits the fact that a group of multi-functional genes may display similarity over a fraction of their interaction profiles if they cooperate in a common process under specific conditions but behave distinctly otherwise. LCD is designed to identify groups of functionally related genes from E-MAP data in a manner that allows individual genes to be assigned to more than one functional group. This enables investigation of the pleiotropic nature of gene function, a goal that cannot be achieved with hierarchical clustering. The performance of our algorithm is illustrated by applying it to two E-MAP datasets (1, 2) and an E-MAP-like *in silico* dataset (3) for the yeast *S. cerevisiae*. In addition to identifying the majority of the functional modules reported in these studies, our algorithm uncovers many recently documented and novel multi-functional relationships between genes and gene groups.

2. MATERIALS AND METHODS

The LCD algorithm is a customized biclustering method applied here to identify genes belonging to common functional modules. Unlike hierarchical clustering methods, which rely on correlating global interaction profiles, the LCD method utilizes both global and local similarities in the genetic interaction profiles and computes a fitness score gauging the extent of profile similarity among genes in a functional module. This score is a product of two terms, a term measuring the signal/noise ratio in the module and another term accounting for the size of the module. To efficiently identify biologically meaningful bi-clusters with the best fitness score, LCD uses Simulated Annealing and Taboo search techniques. The statistical significance of a bicluster is established by computing *p*-values based on comparisons with appropriate random controls. In the following, we present the application of the LCD algorithm to three unrelated E-MAP and E-MAP-like datasets for the yeast *S. cerevisiae* (1-3).

3. RESULTS

The first dataset represents genetic interactions among 890 yeast metabolic genes determined computationally using flux balance analysis (3). Application of the LCD algorithm to this dataset produced 61 modules in total. Seventeen of these are virtually identical to the 'monochromatically interacting' modules derived previously using the Prism algorithm (3), and the genes in these modules display pathway-specific functional annotations. Moreover, our procedure clustered these pathway-specific modules into 21 higher level modules containing genes with related functions. For instance, the 'ATP synthase' module (14 genes) and the 'Electron transport system, complex IV' module (13 genes) are regrouped into a 27-gene module. This is clearly consistent with the fact that electron transport is tightly coupled with ATP synthesis in the mitochondrial inner membrane.

In the second analysis, the LCD algorithm was applied to the experimentally derived E-MAP data for 754 alleles of 743 *S. cerevisiae* genes involved in various aspects of chromosome biology (1). The LCD algorithm identifies 298 modules in this dataset. Among them, there are 28 known protein complexes, 2 known epistatic groups, and 4 sub-complexes. In addition, 4 complexes and 1 sub-complex are also found with a slightly lower fitness score cutoff. The fact that many known protein complexes are recapitulated as genetic modules is a good indication that these modules are biologically meaningful. Several of these complexes or components thereof are further grouped into 151 higher level epistatic modules. The novel view afforded by the LCD algorithm reveals that this grouping is again combinatorial, indicating that complexes may participate in multiple cellular processes, as in the example illustrated in **Fig. 1.**

The third analyzed E-MAP comprises the genetic interactions among 424 *S. cerevisiae* genes involved in the early secretory and protein maturation pathways (2). The resulting network is particularly dense as all the genes belong to intimately linked pathways. Again, the 511 epistatic modules detected in this dataset by our algorithm are enriched in genes sharing identical functional categories and frequently recover known complexes or linear pathways. We found that Spf1, a P-type ATPase located in the endoplasmic reticulum membrane, is linked with genes of diverse functions in 15 2-gene clusters. Rather rewardingly, evidence for 6 of the 15 predicted links could be retrieved from the literature *a posteriori*.

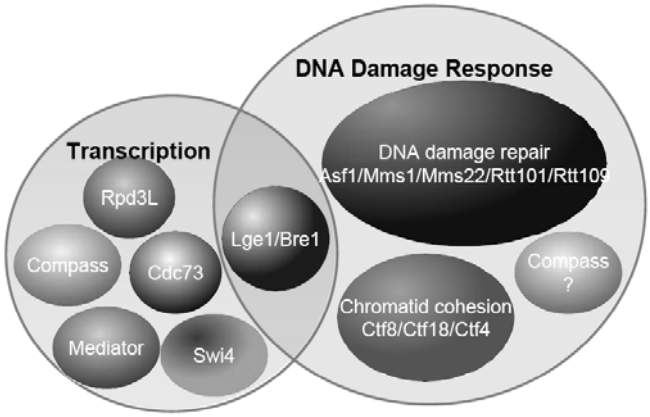

Figure 1. Distinct clusters reveal multiple roles of the Lge1/Bre1 complex. *Results of our biclustering analysis confirm existing experimental evidence implicating the Lge1/Bre1 complex in both transcription and the DNA damage response. The role of the Compass complex in the latter, suggested by our analysis, has not been reported so far.*

4. DISCUSSION

To our knowledge, this is the first attempt in analyzing quantitative genetic interaction data (as opposed to binary data) using biclustering. One of the major advantages of our biclustering approach is that it allows for the grouping of one gene into multiple clusters, making it particularly well suited for uncovering the pleiotropic functions of genes (4). The second advantage of the biclustering approach is its ability to reveal functional links between complexes or pathways. The common function of these complexes may be known in some cases. However, in many cases, complexes that perform diverse roles are found in a cluster, suggesting new roles for known complexes. Many such clusters lack experimental support at the present time, and our findings hence predict many new functional links that would be worthwhile to investigate further.

5. REFERENCES

1. Collins SR, Miller KM, Maas NL, et al. (2007) Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446:** 806-810.
2. Schuldiner M, Collins SR, Thompson NJ, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123:** 507-519.
3. Segre D, Deluna A, Church GM, Kishony R (2005) Modular epistasis in yeast metabolism. *Nature genetics* **37:** 77-83.
4. Dudley AM, Janse DM, Tanay A, Shamir R, Church GM (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. *Molecular systems biology* **1:** 2005.0001.